

RECEIVED  
CENTRAL FAX CENTER

JUN 14 2006

**FACSIMILE COVER PAGE  
(THIS COVER PAGE + 7 PAGES)**

**Today's Date: June 14, 2006**

**To: Examiner L. Ries, Art Unit: 2176**

**FAX: (571) 273-8300/-4095**

**From: Frederick E. Cooperrider #36,769  
McGinn Intellectual Property Law Group, PLLC  
Ph: (703) 761-2377**

**In re Application of Krueken, et al**

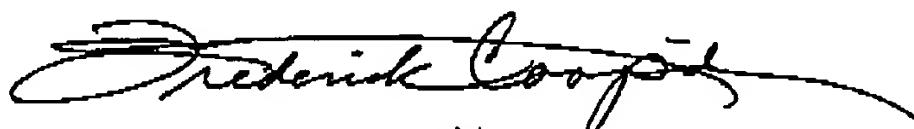
**Serial No.: 09/848,430**

**For: AN EFFICIENT STORAGE MECHANISM FOR REPRESENTING TERM  
OCCURRENCE IN UNSTRUCTURED TEXT DOCUMENTS**

**Contents: Corrected Claims Appendix for Appellants' Brief on Appeal (7 pages), as discussed  
and required on pages 2 and 3 of the Order Remanding Undocketed Appeal to  
Examiner, mailed on June 13, 2006**

**CERTIFICATION OF TRANSMISSION**

I certify that I transmitted via facsimile to (571) 273-8300 this Corrected Claims  
Appendix for Appellants' Brief to the USPTO, plus a copy to 571-273-4095 directly to Examiner  
L. Ries, on June 14, 2006.



Frederick E. Cooperrider  
Reg. No. 36,769

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

RECEIVED  
CENTRAL FAX CENTER  
JUN 14 2006

### **VIII. CLAIMS APPENDIX**

The following claims are presented for review, as reflected upon entry of the Amendment Under 37 C.F.R. §1.111 filed on 9/13/04. It is noted by Appellants that the listing of claims in the Amendment Under 37 C.F.R. §1.111 filed on 4/21/05 have typographical errors in the dependencies listed for claims 2, 3, 6, 7, and 10, as well as mistakes in the status of these claims, as noted in the Order Returning Undocketed Appeal to Examiner, mailed to Appellants on June 13, 2006. These errors are explained as typographical errors made by the preparer of that Amendment filed on 4/21/05 when a set of claim amendments were initially prepared for the Amendment to be filed and then it was decided to revert to the claims as previously presented in the 9/13/04 Amendment. The changes to claims 13 and 15 that were made in the 4/21/05 Amendment are not considered particularly significant to the issues of this Appeal.

1. (Previously presented) A method of converting a document corpus containing an ordered plurality of documents into a compact representation in memory of occurrence data, said method comprising:

developing a first vector for said entire document corpus, said first vector being a listing of integers corresponding to terms in said documents such that each said document in said document corpus is sequentially represented in said listing.

2. (Previously presented) The method of claim 19, further comprising:

developing a third vector for said entire document corpus, said third vector comprising a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor.

3. (Original claim) The method of claim 1, further comprising:

rearranging, in said first vector, an order of said unique integers within the data for each said document so that all identical unique integers are adjacent.

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

4. (Original claim) The method of claim 2, wherein said normalization factor is calculated as:

$NF = 1/(\sum x_i^2)^{1/2}$ , where  $x_i$  is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

5. (Previously presented) A method of converting, organizing, and representing in a computer memory a document corpus containing an ordered plurality of documents, said method comprising:

for said document corpus, taking in sequence each said ordered document and developing a first uninterrupted listing of integers to correspond to an occurrence of terms in the document corpus.

6. (Previously presented) The method of claim 21, further comprising:

developing a third uninterrupted listing for said entire document corpus, said third uninterrupted listing containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

7. (Original claim) The method of claim 5, further comprising:

for each said document in said document corpus, rearranging said unique integers so that any identical integers are adjacent.

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

8. (Original claim) The method of claim 6, wherein said normalization factor is calculated as:

$NF = 1/(\sum x_i^2)^{1/2}$ , where  $x_i$  is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

9. (Previously presented) An apparatus for organizing and representing in a computer memory a document corpus containing an ordered plurality of documents, said apparatus comprising:

an integer determining module receiving in sequence each said ordered document of said document corpus and developing a first uninterrupted listing of said unique integers to correspond to an occurrence of terms in the document corpus.

10. (Previously presented) The apparatus of claim 23, further comprising:

a normalizer developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

11. (Original claim) The apparatus of claim 9, further comprising:

a rearranger rearranging said unique integers so that any identical integers for each said document in said document corpus are adjacent.

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

12. (Original claim) The apparatus of claim 10, wherein said normalizer calculates said normalization factor as:

$NF = 1/(\sum x_i^2)^{1/2}$ , where  $x_i$  is the number of occurrences of a specific term in said document, so that NF represents the reciprocal of the square root of the sum of squares of all term occurrences in said document.

13. (Previously presented) A signal-bearing medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method to organize and represent in a computer memory a document corpus containing an ordered plurality of documents, said method comprising:

developing a first uninterrupted listing of said unique integers to correspond to the occurrence of said dictionary terms in the document corpus.

14. (Previously presented) The signal-bearing medium of claim 25, wherein said method further comprises:

developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

15. (Original claim) A data converter for organizing and representing in a computer memory a document corpus containing an ordered plurality of documents, for use by a data mining applications program requiring occurrence-of-terms data, said representation to be

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

based on terms in a dictionary previously developed for said document corpus and wherein each said term in said dictionary has associated therewith a corresponding unique integer, said data converter comprising:

means for developing a first uninterrupted listing of said unique integers to correspond to the occurrence of said dictionary terms in the document corpus and; and

means for developing a second uninterrupted listing for said entire document corpus containing in sequence the location of each corresponding document in said first uninterrupted listing, wherein said first listing and said second listing are provided as input data for said data mining applications program.

16. (Original claim) The data converter of claim 15, further comprising:

means for developing a third uninterrupted listing for said entire document corpus, containing a sequential listing of floating point multipliers, each said floating point multiplier representing a document normalization factor for a corresponding document in said document corpus.

17. (Original claim) The data converter of claim 15, further comprising:

means for rearranging said unique integers so that any identical integers for each said document in said document corpus are adjacent.

18. (Previously presented) The method of claim 1, further comprising:

developing a dictionary comprising said terms contained in said document corpus; and

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers being said integers comprising said first vector.

19. (Previously presented) The method of claim 1, further comprising:

developing a second vector for said entire document corpus, said second vector indicating the location of each said document's representation in said first vector.

20. (Previously presented) The method of claim 5, further comprising:

developing a dictionary comprising terms contained in said document corpus; and  
associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers used in said first uninterrupted listing.

21. (Previously presented) The method of claim 5, further comprising:

developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.

22. (Previously presented) The apparatus of claim 9, further comprising:

a dictionary developing module to develop a dictionary of terms contained in said document corpus, each said term being associated with a corresponding unique integer.

S/N 09/848,430 - Appellants' Brief on Appeal  
Docket: ARC920000023US1

23. (Previously presented) The apparatus of claim 9, further comprising:

a locator module developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.

24. (Previously presented) The signal-bearing medium of claim 13, wherein said method further comprises:

developing a dictionary comprising terms contained in said document corpus; and associating, with each said dictionary term, an integer to be uniquely corresponding to said dictionary term, said uniquely corresponding integers used in said first uninterrupted listing.

25. (Previously presented) The signal-bearing medium of claim 13, wherein said method further comprises:

developing a second uninterrupted listing for said entire document corpus, said second uninterrupted listing containing, in sequence, the location of each corresponding document in said first uninterrupted listing.